Modelling practical placement of trainee teachers to schools

Katarína Cechlárová[a], Tamás Fleiner[b], David Manlove[c], Iain McBride[c] and Eva Potpinková[a]

[a]Institute of Mathematics, Faculty of Science, P.J. Šafárik University, Jesenná 5, 040 01 Košice, Slovakia

katarina.cechlarova@upjs.sk, eva.potpinkova@student.upjs.sk,

[b]Department of Computer Science and Information Theory, Budapest University of Technology and Economics, Magyar tudósok körútja 2, H-1117 Budapest, Hungary and MTA-ELTE Egerváry Research Group

fleiner@cs.bme.hu,

[c]School of Computing Science, Sir Alwyn Williams Building, University of Glasgow, Glasgow, G12 8QQ, UK

David.Manlove@glasgow.ac.uk, i.mcbride.1@research.gla.ac.uk

Abstract

Several countries successfully use centralized matching schemes for assigning students to study places or fresh graduates to their first positions. In this paper we explore the computational aspects of a possible similar scheme for assigning trainee teachers to schools. Our model is motivated by the situation characteristic for Slovak and Czech education system where each pre-service teacher specializes in two subjects. We show that if the two subjects can be performed independently, then a feasible assignment can be found efficiently by employing network flow techniques, while the requirement to perform both subjects at the same school leads to intractable problems even under several strict restrictions concerning the total number of subjects, partial capacities of schools and the number of acceptable schools each teacher is allowed to list. Finally, we report on an integer programming model for solving the teachers assignment problem and the results of its application to real data.

Keywords: assignment of students, bipartite matching, algorithm, NP-completeness

1 Introduction

The traditional study of teachers-to-be in Slovakia involves the specialization of each student in two subjects, e.g. Mathematics and Physics, Chemistry and Biology, Slovak language and English etc. In addition to the study of various topics of these subjects, principles of Pedagogics and Psychology, each curriculum contains a practical placement in a real school several times during the study. During each placement students visit classes and also teach themselves, always under supervision of an experienced and qualified teacher approved by the university for taking this responsibility. Students might try to find suitable schools and supervising teachers by themselves, but to ensure the quality of such a placement, each faculty usually provides a list of such schools and teachers, and students are assigned to them by the faculty staff.

The assignment task is not easy and as the people responsible for practical placements confirmed to the authors of this paper, it usually takes several days and many iterations to assign each student. In an ideal case, everybody should be assigned to a school in the town where the university is located and use each placement run for a different type of school (an upper elementary, practical secondary or a grammar school). However, even if the number of approved supervising teachers is sufficient for the current number of students to be placed, this is not always possible, as the structure of available places might not be suitable, not all schools provide supervisors for all subjects, or they may not have enough classes to accept several students for a particular subject. Hence some other schools outside the site of the university could be used, but in this case it should be taken into account that it may be infeasible for a student (for economic reasons) to commute to a very distant school.

The aim of this paper is to model the trainee teachers matching problem and to study its computational complexity. It turns out that if it is the case that during one practical placement the student teaches one subject, and during another he/she teaches the second subject (this is the way the practical placement are organized e.g. at Comenius University in Bratislava or at Charles University in Prague) the matching problem can be solved efficiently by emplyoing network flow techniques. However, other universities (e.g. the one where the Slovak authors if this paper are affiliated) require that each student teaches both subjects at one school during the same placement. For this case, we propose efficient algorithms that find a matching for the maximum possible number of students if there are altogether only two specialization subjects, or there are three subjects but each school can accept at most one student for each subject (irrespectively of her other specialization). Interestingly, if for each student at most two schools are acceptable and all partial capacities are at most one, it is possible to decide whether a full assignment (i.e., one that matches all students) exists in polynomial time, but the problem of matching the maximum number of students becomes intractable. The existence of full assignment is also NP-complete if (i) there are only three subjects and schools may have capacity two in one of its subjects, or if (ii) there are four subjects and each school has capacity at most one in each subject. Moreover, this problem is also NP-complete if each school is acceptable for each student, but now without the restriction on the number of subjects.

In the light of the intractability results presented in Section 4, in Section 5 we propose an integer linear program for the teachers assignment problem and applied it to real data. We report on the results of this trial in Section 6.

2 Related work

The classical problems of combinatorial optimization like the maximum cardinality bipartite matching problem, assignment problem, or flow problem have been successfully applied to a range of variants of manpower allocation problems (see e.g. applications reviewed in [4], Chapter 12). Practical situations also lead to some NP-complete variants [10].

Recently, a lot of attention has been attracted by several large-scale centralized allocation schemes used for assigning pupils to public schools in Boston and New York [1, 2], allocationg graduates of medical schools to their first jobs in hospitals in the USA [14], university applicants to study places in Hungary [5] etc.

In such schemes, it is typically the case that the applicants rank in order of preference the entities on the other side of the market (i.e., schools, hospitals or universities) that they find acceptable. In addition, there may be preference lists over applicants belonging to the members of the other side of the market. For an overview of other applications and of various models involving ordinal preferences, together with their computational complexity, the reader is advised to consult the recently published monograph by Manlove [12] or the comprehensive web page containing a description of matching practices for various levels of education and for entry-level labour markets in many European countries.

When considering matching problems involving ordinal preferences, the model that most closely resembles our problem is the Hospitals / Residents problem with Couples [6]. This underpins the problem of assigning medical graduates to their first hospital posts on the basis of two-sided preferences, when couples (i.e., pairs of applicants) are able to submit joint preferences over pairs of hospitals, typically in order to be matched to geographically close positions. A special case of this problem occurs when both members of the couple refuse to be separated and insist on being allocated to the same hospital [13]. Another case is the Scottish scheme for medical students that have to be assigned to two training units (medical and surgical one) and these two assignments have to be allocated to two different semesters [11].

In our problem, the applicants do not have ordinal preferences over schools; they simply list an acceptable set of schools and are effectively indifferent between them. Moreover, the schools do not have preferences over applicants. An existing problem in the literature that most closely resembles this case is called Matching with Couples [7], which is essentially the same as the Hospitals / Residents problem with Couples where agents are indifferent between all members of their preference lists. However we emphasize that our model differs from all of those discussed so far due to the applicants' specialization in two subjects and the schools being allowed to have different capacities for different subjects.

3 Definitions and preliminary observations

An instance J of the Teachers Assignment problem, TAP for short, involves a set A of n applicants (students, trainee teachers), a set S of m schools and a set P of k subjects. For ease of exposition, elements of the set P will sometimes be referred to by letters like M, F, I or B, to remind of real subjects taught at schools, like Mathematics, Physics, Informatics or Biology etc.

Each applicant $a \in A$ is characterized by a pair of different subjects $\mathbf{p}(a) = \{p_1(a), p_2(a)\} \subseteq P$. Sometimes we shall also say that a particular applicant is of type MF, MB, or IB, etc. The set of applicants whose specialization involves a subject $p \in P$ will be denoted by A_p. We also suppose that each applicant a provides a list $S(a)$ of *acceptable* schools, i.e. schools to which he/she is willing to go.

Each school $s \in S$ has a certain capacity for each subject, the vector of capacities of school s will be $\mathbf{c}(s) = (c_1(s), \ldots, c_{|P|}(s)) \in \mathbb{N}^{|P|}$. An entry of $\mathbf{c}(s)$ will be referred to as a *partial capacity* of school s. Here, $c_p(s)$ is the maximum number of applicants whose specialization involves subject p that school s is able to accept. Again, we shall sometimes write $c_M(s), c_I(s)$ etc.

Let us first deal with the case when trainee teachers practice their two subjects separately in two different periods (say first during one semester and then during the second semester of an academic year). Now they can be assigned to two different schools and the main challenge here is to decide the order of subjects for each trainee teacher so as to be able to use the available capacities. For brevity, let us call a student-subject pair a *half-student*.

Let us construct a network $N = (V, E, c)$ where the set of vertices is $V = \{a, a_p, a_q; a \in A, \mathbf{p}(a) = \{p, q\}, \} \cup \{s_p, s \in S, p \in P, c_p(s) \neq 0\} \cup \{r, t\}$. This means, there is a source r, sink t, three vertices for each applicant: a simple one and two for each subject of applicant's specialization. For each school we have one vertex for each subject in which it provides at least one place.

There is an arc ra for each $a \in A$ with capacity $c(ra) = 2$ and two arcs aa_p, aa_q with capacity 1 for the two subjects $p, q \in \mathbf{p}(a)$. There is an arc with unit capacity from a_p

to s_p if school s is acceptable for applicant a. For each vertex s_p we add an arc s_pt with capacity $c(s_pt) = 2c_p(s)$.

Theorem 1 *There is an integer flow f of size K in N if and only if it is possible to place K half-students during an academic year.*

Clearly, all students can be placed for their both subjects if and only if the size of the maximum flow in N is $2n$. As all the capacities in N are integer, the flow along each arc is integral and its interpretation is clear: applicant a will perform her specialization subject p at school s if and only if there is a unit flow along the arc a_ps_p.

To divide the half-students into two periods, we use the following trick. Replace the capacity $c(e)$ of each arc e of N by $\lceil c(e)/2\rceil$ and take the flow $g(e) = f(e)/2$. The halved flow is no longer integer, but as the new capacities are, we use the following assertion, known as Integrality lemma [15].

Lemma 2 (Integrality Lemma.) *Let $D = (V, E, c)$ be a network and f a flow of value $K \in \mathbb{Z}$. Then there exists an integer flow g of value K with $\lfloor f(e)\rfloor \le g(e) \le \lceil f(e)\rceil$ for each arc e.*

It is clear that at most one of the arcs $e \in \{aa_p, aa_q\}$ for each applicant has $g(e) = 1$. Let us say that applicant a will practice her subject p in period one at school s such that $g(aa_p) = g(a_ps_p) = 1$ and subject q in period one at school s such that $g(aa_p) = g(a_ps_p) = 0$ and simultaneously $f(aa_p) = f(a_ps_p) = 1$.

By this, the case with separated subjects is completely solved and from now on we deal only with the case when both subjects are practiced simultaneously at the same school.

An *assignment* $\mathcal{M}$ is a subset of $A \times S$ such that each applicant $a \in A$ is a member of at most one pair in $\mathcal{M}$. We shall write $\mathcal{M}(a) = s$ if $(a, s) \in \mathcal{M}$ and say that applicant a is *assigned* (to school s); if there is no such school, applicant a is *unassigned*. The set of applicants assigned to a school s will be denoted by $\mathcal{M}(s) = \{a \in A; (a, s) \in \mathcal{M}\}$. We shall also denote by $\mathcal{M}_p(s)$ the set of applicants assigned to s whose specialization includes subject p and by $\mathcal{M}_{p,r}(s)$ the set of applicants assigned to s whose specialization is exactly the pair $\{p, r\}$. More precisely,

$$\mathcal{M}_p(s) = \{a \in A; (a, s) \in \mathcal{M} \ \&\ p \in \mathbf{p}(a)\}$$

and

$$\mathcal{M}_{p,r}(s) = \{a \in A; (a, s) \in \mathcal{M} \ \&\ \{p, r\} = \mathbf{p}(a)\}.$$

An assignment $\mathcal{M}$ is *feasible* if $\mathcal{M}(a) \in S(a)$ for each $a \in A$ and $|\mathcal{M}_p(s)| \le c_p(s)$ for each school s and each subject p.

Example. Suppose there are 3 subjects M,F and I and four applicants: a_1 of type IF, a_2 of type MF and a_3, a_4 of type MI. There are two schools s_1, s_2 with $c_M(s_1) = 1$, $c_F(s_1) = c_I(s_1) = 2$ and $c_M(s_2) = 2$, $c_F(s_2) = c_I(s_2) = 1$. Both schools are acceptable for all applicants.

Here it is possible to assign all applicants, namely $\mathcal{M}(a_1) = \mathcal{M}(a_3) = s_1$ and $\mathcal{M}(a_2) = \mathcal{M}(a_4) = s_2$. However, suppose that applicant a_1 is assigned to school s_2. Then the three applicants a_2, a_3 and a_4 compete to the unique place in Mathematics at school s_1, hence at most one of them can be assigned.

This shows that there may exist maximal matchings (i.e., such that no additional applicant can be placed) with cardinality equal to half the cardinality of maximum matching.

MAX-TAP denotes the problem to decide, given an instance J of TAP and an integer ℓ, whether a feasible assignment exists that assignes at least ℓ applicants. A special case of MAX-TAP asking for an assignment that leaves no student unassigned will be denoted by FULL-TAP. In the following section we explore the computational complexity of several special cases of this problem.

4 Computational complexity

In this section we show that MAX-TAP is easy in some very restricted cases (as far as the number of subjects or the size of preference lists is concerned). Later we show that in situations that usually occur in practice, the problem is intractable.

Theorem 3 MAX-TAP *is solvable in polynomial time in each of the following cases:*

(i) $|P| = 2$*;*

(ii) $|P| = 3$ *and no partial capacity of a school exceeds* 1.

Proof. For *(i)* it suffices to realize that all applicants are essentially equivalent and a school with partial capacities c_1 and c_2 can admit at most $c = \min\{c_1, c_2\}$ students. Hence MAX-TAP reduces to the classical bipartite *b*-matching problem that can be solved in polynomial time by any well-known algorithm, see e.g. [4].

Similarly, in *(ii)* each school can admit at most one applicant, so MAX-TAP is reduced to the simple maximum cardinality bipartite matching problem, again solvable in polynomial time. ■

In the following two assertions let us denote by TAP(α, β) the set of instances of TAP in which each applicant is allowed to list at most α acceptable schools and all partial capacities are at most β. Hence FULL-TAP(α, β) and MAX-TAP(α, β) denote the problems defined above restricted to instances in TAP(α, β).

Theorem 4 FULL-TAP$(2, 1)$ *is solvable in polynomial time for* $|P|$ *arbitrary.*

Proof. Let us proceed in the following way. In the first phase we deal with applicants that list a school that does not have enough capacity for both specialization subjects. Such schools can be removed from their lists. If we get some applicants with empty lists, then the particular instance of FULL-TAP is clearly insolvable. Otherwise, if the list of an applicant contains only one school (let us call these applicants *spoiled*), to get a full assignment, he/she must be assigned to that particular school. This, however, decreases the respective partial capacities of the school involved and new spoiled applicants can emerge. If, in this first phase we are not able to place all spoiled applicants, no full matching exists; otherwise we continue with the second phase with the partial capacities reduced accordingly. (It is easy to see that the first phase can be performed in polynomial time.)

The obtained *cannonical* FULL-TAP instance J has $|S(a)| = 2$ for each $a \in A$. Let us denote $S(a_i) = \{s_i^1, s_i^2\}$ and introduce a boolean variable x_i for each applicant a_i with the following interpretation: if x_i is TRUE, we shall say that a_i is assigned to school s_i^1; if x_i is FALSE, we say that a_i is assigned to school s_i^2. Now create a boolean formula $B(J)$ in the following way. For each pair of applicants a_i, a_j whose specialization involves at least one common subject and for each school $s \in S(a_i) \cap S(a_j)$ we create a clause $C_{i,j,s}$ as follows:

- if $s = s_i^1$ and $s = s_j^1$ then $C_{i,j,s} = \bar{x}_i + \bar{x}_j$;
- if $s = s_i^1$ and $s = s_j^2$ then $C_{i,j,s} = \bar{x}_i + x_j$;

- if $s = s_i^2$ and $s = s_j^1$ then $C_{i,j,s} = x_i + \bar{x_j}$;
- if $s = s_i^2$ and $s = s_j^2$ then $C_{i,j,s} = x_i + x_j$.

Clause $C_{i,j,s}$ ensures that a_i and a_j do not both occupy the only place for their common subject at school s. Formula $B(J)$ is then the conjuction of clauses $C_{i,j,s}$ for all triples a_i, a_j, s as described above. It is easy to see that $B(J)$ is satisfiable if and only if a full assignment for J exists (remember, we assume that J is cannonical). $B(J)$ is a boolean formula in conjunctive normal form and since each clause contains just two literals, its satisfiability can be decided in polynomial time. This concludes that FULL-TAP(2,1) is polynomially solvable. ■

By contrast, maximizing the number of assigned applicants is difficult.

Theorem 5 MAX-TAP$(2,1)$ *is NP-complete.*

Proof. MAX-TAP$(2,1)$ clearly belongs to NP. To prove completeness, we give a polynomial reduction from the 3-coloring problem restricted to graphs G with maximum degree $\Delta(G)$ bounded by 4 (see [9], problem GT4).

So let us assume that $G = (V, E)$ is a simple graph with $\Delta(G) \leq 4$. For G we construct in polynomial time an instance J of TAP$(2,1)$ and an integer ℓ in such a way that G admits a proper vertex coloring by three colors (let us denote them by r, w and b, i.e., red, white and blue) if and only if J admits a feasible assignment assigning at least ℓ applicants.

By Vizing's theorem, the edges of G can be colored by five colors, let $c : E \to \{1, 2, 3, 4, 5\}$ denote a proper coloring of E, i.e., such that no two edges incident to the same vertex have the same color. This coloring will be used in our construction.

There are 38 subjects in J: two subjects $p(c, x, 1)$ and $p(c, x, 2)$ for each (edge color, vertex color) pair (c, x) (altogether 30 subjects) plus 8 subjects p^1, p^2, p_x^1, p_x^2 for $x \in \{r, w, b\}$ (vertex subjects).

There is one school s_v for each vertex $v \in V$. School s_v provides one place for each vertex subject and one place for each subject $p(c, x, j)$, where $x \in \{r, w, b\}$, $j = 1, 2$ and c are colors of edges incident to v. This means, there are altogether $8|V| + \sum_{v \in V} 3 \cdot 2 \cdot deg(v) = 8|V| + 12|E|$ places at schools, capable of admitting $\ell = 4|V| + 6|E|$ applicants.

For each edge $e = \{u, v\} \in E$ define three applicants: $a(e, r), a(e, w)$ and $a(e, b)$ and call them the red, white and blue applicant of edge e, respectively. Their subjects are $p(c(e), j, 1)$ and $p(c(e), x, 2)$ where $c(e)$ is the color of edge e and x is the color of the applicant, i.e., either red, white or blue. The acceptable schools of these three applicants are just s_v and s_u.

To define further applicants, let us consider for each vertex $v \in V$ an auxiliary graph G_v, illustrated in Figure 1.

Suppose that v is incident with edges $e_1, e_2, \ldots, e_t$, where $t \leq 4$. G_v consists of three edge-disjoint paths, each corresponding to one of the vertex colors r, w, b. Vertices of G_v represent subjects offered at school s_v, the subjects along one path corresponding to $x \in \{r, w, b\}$ are in this order

$$p^1, p_x^1, p(c(e_1), x, 1), p(c(e_1), x, 2), p(c(e_2), x, 1), p(c(e_2), x, 2), \\ \ldots, p(c(e_t), x, 1), p(c(e_t), x, 2), p_x^2, p_2.$$

In Figure 1, thick edges correspond to already defined edge applicants. Notice that these applicants appear also in graphs G_u for each u adjacent to v. All the other edges in G_v define new applicants (by the corresponding pair of subjects) called *vertex applicants* whose the only acceptable school is s_v. Hence in J we have $3|E|$ edge applicants and

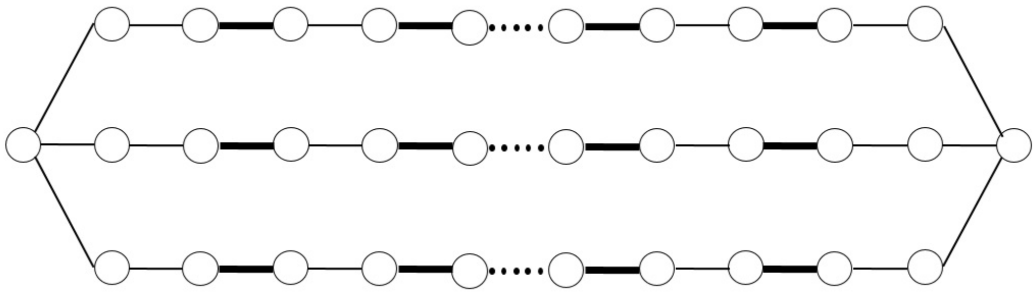

Figure 1: Auxiliary graph

$3(2deg(v)+3)$ vertex applicants for each vertex $v \in V$, hence altogether $9|V|+15|E|$ applicants.

Notice that since we used a proper edge coloring of G, no two subjects defined for vertices of G_v are the same. Further, G_v has exactly three perfect matchings: M_r, M_w and M_b, where M_x, $x \in \{r, w, b\}$ matches all the edge students corresponding to all edges incident to v of color x (they are on one of the paths) and all vertex students on the other two paths. Each perfect matching of G_v defines an assignment of applicants that occupy all places offered by school s_v.

Now suppose that G admits a proper coloring of its vertices by three colors. For each $v \in V$, let us choose in G_v matching M_r, M_w or M_b, according to whether v is colored by red, white or blue. As two adjacent vertices received a different color, this is a feasible matching assigning $\ell = 4|V| + 6|E|$ applicants.

Conversely, let J admit a matching that assigns $k = 4|V| + 6|E|$ applicants. Such an assignment has to fill all the places at each school, hence the applicants assigned to school s_v must correspond to a perfect matching in G_v, for each $v \in V$. Let us color vertex v by one of the colors red, white or blue, if G_v is matched according to M_r, M_w or M_b. Now consider two vertices v and u joined by edge e and suppose that they are colored by the same color (say red). However, by the definition of the coloring, the red applicant $a(e, r)$ corresponding to edge e is assigned to s_v as well as to s_u, a contradiction. ∎

Let us also remark here that the previous theorem implies that the problem to decide whether there exists a feasible assignment using all the available places at schools is also NP-complete.

The starting known NP-complete problem for the following theorem is 3-dimensional matching, 3DM in brief ([9], problem SP1). An instance of 3DM contains three disjoint sets U, V and W, all of cardinality n, and a set of triples $\mathcal{T} \subseteq U \times V \times W$. The question is whether there exists a perfect matching, i.e. a subset $\mathcal{N} \subseteq \mathcal{T}$ of cardinality n that covers all elements of $U \cup V \cup W$. We shall use the NP-complete restriction of 3DM to such instances where no element occurs in more than 3 triples in $\mathcal{T}$.

Theorem 6 FULL-TAP *is NP-complete already when* $|S(a)| \leq 3$ *and*

(*i*) $|P| = 3$ *and no partial capacity of a school exceeds* 2*; or*

(*ii*) $|P| = 4$ *and no partial capacity of a school exceeds* 1.

Proof. For case (*i*), given an instance $J = (U, V, W, \mathcal{T})$ of 3DM, we construct an instance J' of TAP with 3 subjects (say M,F and I) and $c_M(s) = 2$, $c_F(s) = c_I(s) = 1$ for each school.

For each triple $t \in \mathcal{T}$ we create a school s_t. For each $z \in U \cup V \cup W$ let $\mathcal{T}_z$ be the set of triples in $\mathcal{T}$ containing z and $\ell_z = |\mathcal{T}_z|$. For each $u \in U$ we create applicants

$a_u^1, a_u^2, \dots, a_u^{\ell_u - 1}$, each of type IF; their set will be denoted by A_u. For each $v \in V$ we create an applicant a_v of type MI and for each $w \in W$ an applicant a_w of type MF. For each applicant corresponding to an element $z \in U \cup V \cup W$, acceptable schools are those that correspond to triples in $\mathcal{T}_z$.

Suppose that the 3DM instance J has a perfect matching $\mathcal{N} \subseteq \mathcal{T}$. We assign each applicant in J' to an acceptable school so that the capacity of no school in no subject will be exceeded.

For each $t = (u, v, w) \in \mathcal{N}$ we assign to school s_t applicants a_v and a_w. For each $u \in U$ there are $\ell_u - 1$ triples $t \in \mathcal{T} \backslash \mathcal{N}$ containing u, so to the corresponding schools we assign applicants $a_u^1, a_u^2, \dots, a_u^{\ell_u - 1}$. It is easy to see that each applicant is assigned to an acceptable school and that the defined assignment obeys all capacities.

Conversely suppose that there exists a full feasible assignment $\mathcal{M}$. Let $S_{\mathcal{N}}$ be the set of schools to which two applicants are assigned in $\mathcal{M}$ and let $\mathcal{N} \subseteq \mathcal{T}$ be the set of corresponding triples. By the construction, if $s_t \in S_{\mathcal{N}}$ and $t = (u, v, w)$ then the assigned applicants are a_v and a_w. Clearly, for two different schools in $S_{\mathcal{N}}$ these two applicants are different and so also any two different triples in $\mathcal{N}$ differ in their elements from V and W. It remains to show that if $t, t' \in \mathcal{N}$ are different then their corresponding elements from U are also different.

To get a contradiction, suppose that some element $u \in U$ belongs to at least two different triples $t, t' \in \mathcal{N}$. Notice that the only acceptable schools for $\ell_u - 1$ applicants of the set A_u are the ℓ_u schools s_t for $t \in \mathcal{T}_u$. If two different schools $s_t, s_{t'}$ belong to $S_{\mathcal{N}}$ then the number of schools that have enough capacity for $\ell_u - 1$ applicants in A_u and are acceptable for them is at most $\ell_u - 2$. This is a contradiction with the assumption that $\mathcal{M}$ is a full assignment.

The proof for (i) can easily be modified for (ii) by making the following changes:

- The set of subjects is M,F,I,B,
- each school s has $c_M(s) = c_F(s) = c_I(s) = c_B(s) = 1$;
- for each $v \in V$ the type of applicant a_v is MF;
- for each $w \in W$ the type of applicant a_v is IB;
- for each $u \in U$ contained in ℓ_u triples in $\mathcal{T}$ there are $\ell_u - 1$ applicants of type MI and $\ell_u - 1$ applicants of type FB.

The acceptability is defined in the same way according to the structure of $\mathcal{T}$ and the rest of the proof is analogical. ■

The last assertion in this section concerns the case when each school is acceptable for each applicant (e.g. if each school is in the town where the university is located). However, we need a great number of subjects and the computational complexity of the problem with bounded number of subjects remains open.

Theorem 7 MAX-TAP *is NP-complete, even in the case when each school is acceptable for each applicant and no partial capacity exceeds 2.*

Proof. We reduce from a restricted version of SAT. Let (2,2)-E3-SAT denote the problem of deciding, given a Boolean formula B in CNF in which each clause contains exactly 3 literals and for each variable v_j, each of literals v_j and $\bar{v}_j$ appears exactly twice in B, whether B is satisfiable. Berman et al. showed that (2,2)-E3-SAT is NP-complete.

Hence let B be an instance of (2,2)-E3-SAT. Let $V = \{v_1, v_2, \dots, v_n\}$ and $C = \{c_1, c_2, \dots, c_m\}$ be the set of variables and clauses respectively in B. Let us denote the indices of the two clauses that contain literal v_j by $c(v_j^1)$ and $c(v_j^2)$, and the indices of the two clauses that contain literal $\bar{v}_j$ by $c(\bar{v}_j^1)$ and $c(\bar{v}_j^2)$, respectively. For each clause c_i, the symbols $v^1(c_i)$, $v^2(c_i)$ and $v^3(c_i)$ denote the indices of the variables (negated or not) that appear in the first, second and third position in c_i, respectively.

Now we construct an instance J of TAP in the following way.

For each variable v_j, $j = 1, 2, \ldots, n$ there are 3 subjects x_j, y_j^1, y_j^2 and two schools s_j^T, s_j^F. For an easy reference, they will be called the *variable subjects* and *variable schools.* For each clause c_i, $i = 1, 2, \ldots, m$, there is one subject $\hat{p}_i$ and one school $\hat{s}_i$. Hence there are altogether $3n + m$ subjects and $2n + m$ schools. Each partial capacity at each school is at most 2. Table 1 gives for each school the list of subjects with capacity 1 and the (only) subject with capacity 2.

There is one a-applicants for each literal and three b-applicants for each variable. The types of applicants are given in Table 2.

school	subjects with capacity 1	subject with capacity 2
s_j^T	$y_j^1, y_j^2, \hat{p}_{c(v_j^1)}, \hat{p}_{c(v_j^2)}$	x_j
s_j^F	$y_j^1, y_j^2, \hat{p}_{c(\bar{v}_j^1)}, \hat{p}_{c(\bar{v}_j^2)}$	x_j
$\hat{s}_i$	$x_{v^1(c_i)}, x_{v^2(c_i)}, x_{v^3(c_i)}$	$\hat{p}_i$

Table 1: The subjects with capacities 1 and 2 at schools

applicant	subject 1	subject 2	applicant	subject 1	subject 2
$a(v_j^1)$	x_j	$\hat{p}_{c(v_j^1)}$	b_j	y_j^1	y_j^2
$a(v_j^2)$	x_j	$\hat{p}_{c(v_j^2)}$	$\hat{b}_j^1$	x_j	y_j^1
$a(\bar{v}_j^1)$	x_j	$\hat{p}_{c(\bar{v}_j^1)}$	$\hat{b}_j^2$	x_j	y_j^2
$a(\bar{v}_j^2)$	x_j	$\hat{p}_{c(\bar{v}_j^2)}$			

Table 2: Types of applicants

Due to the definitions of subjects it is clear that the a-applicants can only be assigned to schools that correspond either to the variable or to the clause that contains the corresponding literal and the b-applicants related to a variable can only be assigned to schools that correspond to this variable.
For each j, $1 \le j \le n$, let us denote

$$T_j = \{(b_j, s_j^T), (\hat{b}_j^1, s_j^F), (\hat{b}_j^2, s_j^F)\}, \qquad F_j = \{(b_j, s_j^F), (\hat{b}_j^1, s_j^T), (\hat{b}_j^2, s_j^T)\}.$$

Now let f be a satisfying truth assignment of B, we shall define a full assignment $\mathcal{M}$ as follows. If variable v_j is true, add the pairs in T_j to $\mathcal{M}$, if v_j is false, add the pairs in F_j to $\mathcal{M}$. Applicants of type a corresponding to true literals of variable v_j are assigned to school s_j^T, those that correspond to false literals of variable v_j are assigned to the school that corresponds to the clause containing the literal in question. As each clause contains at most two false literals, it is easy to check that each applicant is assigned to a school offering both her subjects and that no partial capacity is exceeded.

Conversely, let $\mathcal{M}$ be a full feasible assignment for J. Then, for each j, the b-applicants are assigned according to T_j or F_j. Let us call the former case the T case and the latter the F case. Define the truth assignment f by setting variable v_j to be true in the T case and false in the F case.

Now take clause c_i. As the capacity of subject $\hat{p}_i$ at school $\hat{s}_i$ is two, at most two a-applicants who listed school $\hat{s}_i$ can be assigned to $\hat{s}_i$. So at least one applicant corresponding to a literal contained in c_i, say a literal of variable v_j, has to be assigned to school s_j^T or s_j^F, according to whether this literal is v_j or $\bar{v}_j$. Now it is easy to see that in accordance with the definition of f, this must be a true literal. Hence B is satisfied by f. ■

5 Integer linear program for TAP

Taking into account the intractability results of the previous section, we now turn our attention to integer linear programming. In this section we present an integer linear programming model for solving the MAX-TAP problem. This formulation allows also for some special features that were encountered in the real data. First, we allow that some applicants need a placement for one subject only. This situation occurs if an applicant has received recognition for one of their subjects for some other activity (e.g. teaching the subject in question in a specialized summer camp, working in a councelling centre, etc.) or if the applicant has failed an exam that is a prerequisite for a particular subject and cannot therefore study that subject before resitting and passing the exam at a later date. We shall also provide a special treatment for cases where students specializing in one subject are allowed to do their practical placement in another (related) subject. This is the case for students of Psychology. Since there are an insufficient number of posts to allocate all of the students who wish to study Psychology these students may be allocated to either Ethics or Civics courses instead.

Let J be an instance of TAP with applicants $A = \{a_1, \ldots, a_n\}$, schools $S = \{s_1 \ldots, s_m\}$ and subjects $P = \{p_1, \ldots, p_k\}$. Let us associate with each applicant $a_i \in A$ a vector v^i of length k such that $v_p^i = 1$ if the specialization of a_i involves subject p and $v_p^i = 0$ otherwise. Thus $v_p^i = 1$ for at most two values of p, for a given i, $(1 \le i \le n)$. Further, each applicant a_i has an ordered list of length $\ell(a_i)$ consisting of acceptable schools $s \in S(a_i)$. Let $pos(a_i, r)$ denote the school at position r in the ordered list of applicant a_i, where $1 \le i \le n$ and $1 \le r \le \ell(a_i)$.

The set of variables will be $X = \{x_{i,r}; 1 \le i \le n; 1 \le r \le \ell(a_i) + 1\}$ with the following interpretation:

$$x_{i,r} = \begin{cases} 1 & \text{if } a_i \text{ is matched to the school in position } r^{th} \text{ in his list} \\ 0 & \text{otherwise} \end{cases}$$

for $r = 1, 2, \ldots, \ell(a)_i$, and

$$x_{i,\ell(a_i)+1} = \begin{cases} 1 & \text{if } a_i \text{ is unmatched} \\ 0 & \text{otherwise.} \end{cases}$$

Let us consider the following linear program

$$\sum_{i=1}^{n} \sum_{r=1}^{\ell(a_i)} x_{i,r} \quad \to \quad max \tag{1}$$

$$\sum_{r=1}^{\ell(a_i)+1} x_{i,r} \quad = \quad 1 \text{ for } i = 1, \ldots, n \tag{2}$$

$$\sum_{i=1}^{n} \sum_{r=1}^{\ell(a_i)} \{x_{i,r}; pos(a_i, r) = s_j; v_p^i = 1\} \quad \le \quad c_p(s_j) \text{ for } j = 1, \ldots, m,\ p = 1, \ldots, k \tag{3}$$

$$x_{i,r} \quad \in \quad \{0, 1\} \tag{4}$$

Constraints (2) ensure that each applicant is matched to exactly one school or is unmatched. Constraints (3) express that the number of applicants assigned to school s_j whose specialization involves subject p does not exceed the partial capacity of subject p at school s_j. The following assertion is obvious.

Theorem 8 *The optimal solution of (1)–(4) corresponds to a solution of* MAX-TAP*.*

Let us now describe how to handle the possibility that there are an unsufficient number of places for applicants whose specialization involves a certain subject (for simplicity, let us suppose that the index of this subject is 1), but it is acceptable to assign these applicants to places of some related subjects (here, again for ease of exposition, let us suppose that these related subjects are indexed by 2 and 3). First, let us denote the set of applicants a_i such that $v_1^i = 1$ and $v_2^i = v_3^i = 0$ by A'. For each $a_i \in A'$ we creat two clones a_{i+n} and a_{i+2n}, such that

$$\begin{aligned}
&v_1^{i+n} = 0; \;\; v_2^{i+n} = 1; \;\; v_3^{i+n} = 0; \\
&v_1^{i+2n} = 0; \; v_2^{i+2n} = 0; \; v_3^{i+2n} = 1; \\
&v_j^{i+n} = v_j^{i+2n} = v_j^i \text{ for each } j > 3.
\end{aligned}$$

The lists of schools both clones of a_i are the same as that of a_i. The constraints which are applied to a_i are applied in similar fashion to a_{i+n} and a_{i+2n}.

Since we require that at most one of the three clones be matched, the unmatched position $\ell(a_i) + 1$ may be 0 for at most one of the three clones. Thus the sum across the 3 unmatched positions must be greater than or equal to 2. Thus ILP (1)–(4) we add for each 'cloned' applicant the following constraint

$$x_{i,\ell(a_i)+1} + x_{i+n,\ell(a_{i+n})+1} + x_{i+2n,\ell(a_{i+2n})+1}/geq2.$$

The clone that is actually matched determines which subject will applicant a_i teach (either subject 1, or subject 2 or subject 3).

6 Description of data

The teachers study at the Faculty of Science, P.J. Šafárik University, Košice, started in 1953. Originally, only four subjects were offered: Mathematics, Physics, Chemistry and Biology. These were joined in 2002 by Informatics and Geography. In reaction to strictly decreasing numbers of student, the Faculty of Science decided to join forces with the newly created Philosophical Faculty (2007) and offer joined study programmes. Currently the number of subjects offered by the Faculty of Science is 6, Philosophical Faculty offers 8 subjects. Practically any combination of two subjects is possible, in our dataset we encountered 33 different pairs of subjects.

In the current academic year there are approximately 500 students studying teachers combinations at both faculties. In one run, between 100 and 150 of them have to be placed. There are 175 schools. The numbers of supervising teachers for various subjects and the numbers of students who applied for the placement in the Spring 2014 run whose specialization contains a given subject are given in Table 3. Recall that when the time comes, some of these students will not participate in the placement or will be allowed to teach only one of their specialization subjects (e.g. if they fail at some prerequisite subjects at the end of the winter semester). Our model is capable of taking this into account.

When looking at Table 3, it seems that the numbers of supervising teachers is sufficient, except for Psychology. The common practice is to assign students of Psychology

Places	Mathematics	Physics	Biology	Chemistry	Informatics	Geography	Slovak	English	German	Latin	Civics	Psychology	Ethics	History
Košice	74	36	50	38	44	31	54	57	35	0	24	2	16	23
Total	288	158	172	142	137	127	243	216	129	3	119	12	80	135
Applicants	13	9	43	21	4	35	31	14	22	1	21	22	12	28

Table 3: Numbers of supervising teachers and applicants for individual subjects

to supervising teachers of Civics or Ethics, as described in the previous section. Then the total number of teachers of these three subjects is sufficient for the total number of students whose specialization involves the three subjects.

With these assumptions, our ILP model found that out of 138 students, the maximum number that could be placed to schools in Košice was 122. When we included also rural schools, this number increased to 137. It was easy to identify the only student who could not be placed: his specialization involved Latin and there was no Latin teacher in a school acceptable for this student. The time needed for the computations was in both cases less than one second.

7 Conclusions and open problems

In the quest for a possible centralized matching scheme the presented intractability results are pessimistic. Still, integer programming formulation proved to be very effective for solving the given problem in a practical context.

The existing extensive literature on matchings and many succesfull existing schemes call for exploring other possible approaches. Students, in addition to expressing acceptability, could be allowed to list the acceptable schools in order of their preference and/or the schools might also be given the right to order students. Then some other criteria for the obtained matching might be considered: Pareto optimality (from the viewpoint of students, see [3]) or stability (introduced by Gale and Shapley [8]).

8 Acknowledgements

This work was supported by VEGA grant 1/0479/12 (Cechlárová, Potpinková), by OTKA K108383 and the ELTE-MTA Egerváry Research Group (Fleiner), by a SICSA Prize PhD Studentship and COST (IC1205) Action on Computational Social Choice (McBride), by Engineering and Physical Sciences Research Council grant EP/K010042/1 (Manlove). The authors also gratefully acknowledge the support of the Operational Program "Education and Research" funded by the European Social Fund, grant "Education at UPJŠ Heading towards Excellent European Universities", ITMS project code: 26110230056.

We would also like to thank Renáta Orosová, Nataša Kocová, Tatiana Bušová, Roman Soták (UPJŠ Košice), Iveta Kohanová (Comenius University Bratislava) and Nad'a Vondrová (Charles University Prague) for detailed explanations of practical aspects of student assignment and for providing our the data.

References

[1] A. Abdulkadiroğlu, P.A. Pathak, and A.E. Roth. The Boston public school match. *American Economic Review*, 95(2):368–371, 2005.

[2] A. Abdulkadiroğlu, P.A. Pathak, and A.E. Roth. The New York City high school match. *American Economic Review*, 95(2):364–367, 2005.

[3] D.J. Abraham, K. Cechlárová, D.F. Manlove, and K. Mehlhorn. Pareto optimality in house allocation problems. In *Proceedings of ISAAC '04: the 15th Annual International Symposium on Algorithms and Computation*, volume 3341 of *Lecture Notes in Computer Science*, pages 3–15. Springer, 2004.

[4] R.K. Ahuja, T.L. Magnanti, and J.B. Orlin. *Network flows: Theory, algorithms, and applications*. Prentice Hall, 1993.

[5] P. Biró, T. Fleiner, R.W. Irving, and D.F. Manlove. The College Admissions problem with lower and common quotas. *Theoretical Computer Science*, 411:3136–3153, 2010.

[6] P. Biró and F. Klijn. Matching with couples: a multidisciplinary survey. *International Game Theory Review*, 15(2), 2013. article number 1340008.

[7] P. Biró and E. McDermid. Matching with sizes (or scheduling with processing set restrictions). *Discrete Applied Mathematics*, 164(1):61–67, 2014.

[8] D. Gale and L.S. Shapley. College admissions and the stability of marriage. *American Mathematical Monthly*, 69:9–15, 1962.

[9] M.R. Garey and D.S. Johnson. *Computers and Intractability*. Freeman, San Francisco, CA., 1979.

[10] A. Hefner and P. Kleinschmidt. A constrained matching problem. *Annals of Operations Research*, 57:135–145, 1995.

[11] R.W. Irving. Matching medical students to pairs of hospitals: a new variation on a well-known theme. In *Proceedings of ESA '98: the 6th European Symposium on Algorithms*, volume 1461 of *Lecture Notes in Computer Science*, pages 381–392. Springer, 1998.

[12] D.F. Manlove. *Algorithmics of Matching Under Preferences*. World Scientific, 2013.

[13] E.J. McDermid and D.F. Manlove. Keeping partners together: Algorithmic results for the hospitals / residents problem with couples. *Journal of Combinatorial Optimization*, 19(3):279–303, 2010.

[14] A.E. Roth. The evolution of the labor market for medical interns and residents: a case study in game theory. *Journal of Political Economy*, 92(6):991–1016, 1984.

[15] A. Schrijver. *Combinatorial Optimization: Polyhedra and Efficiency*, volume 24 of *Algorithms and Combinatorics*. Springer, 2003.